A Survey of Novelty Detection in Encrypted Data

Arsalan, Obaid Sheikh Ahmed, & Muhammad Talha Panhwar

Department of Artificial Intelligence, Sindh Madressatul Islam University
Karachi, Pakistan

Abstract

The increasing use of encrypted network traffic, while essential for user privacy, has created a significant challenge for conventional intrusion detection systems (IDS). Attackers are exploiting this encryption to conceal malicious activities, rendering traditional security tools ineffective. Our project addresses this critical problem by developing a privacy preserving, AI-based system that can accurately detect hidden cyber threats within encrypted data without the need for decryption. This is achieved by focusing on the behavioral and statistical patterns of network flows rather than their content.

Our research methodology involved using two prominent public datasets, CICID 2017 and UNSW-NB15, which were carefully processed for our models. We trained and evaluated a variety of machine learning models, including powerful algorithms like XGBoost and Random Forest, to find the best approach. Furthermore, we designed a novel hybrid model that combines a deep learning autoencoder for intelligent feature extraction with an XGBoost classifier for highly accurate threat classification.

The key findings from our experiments show the effectiveness of our approach. On the CICID 2017 dataset, the XGBoost classifier achieved an impressive accuracy of 98.62%, while the Random Forest model performed exceptionally well on the UNSW-NB15 dataset, reaching an accuracy of 99.23%. Our proposed hybrid model also showed strong performance, achieving a 98.15% accuracy on the CICID 2017 dataset, demonstrating its ability to handle complex data patterns.

In conclusion, our study successfully proves that an AI system can effectively and accurately identify cyber threats from encrypted network traffic by analyzing flow-based features, thereby upholding both security and privacy standards.

Keywords: AI, Cyber Threat, Encrypted Data, Intrusion Detection, Machine Learning, Deep Learning

1. Introduction

Background: The Importance of Data Security in an Interconnected World

In today's highly interconnected world, where every aspect of our lives from personal communications to financial transactions is conducted online, data security has become a paramount concern. The sheer volume and sensitivity of data traversing global networks necessitate robust security measures. A key technology for protecting our online information is **encryption**. By transforming data into an unreadable format, encryption acts as a powerful safeguard against unauthorized access. It is a cornerstone of digital privacy and a fundamental requirement for building trust in the digital ecosystem. However, this very technology, designed to protect us, has also created a new and difficult challenge for cybersecurity professionals.

The Security Challenge of Encrypted Traffic

- **The Rise of Encrypted Threats**: The growing adoption of encryption has created a "dark space" within network traffic. Cyber attackers are increasingly exploiting this anonymity to camouflage malicious activities, such as propagating viruses, launching sophisticated ransomware, and conducting covert data theft. This strategy allows them to bypass conventional security checkpoints by hiding their actions inside what appears to be legitimate, encrypted traffic.
- **The Ineffectiveness of Traditional Systems**: This new wave of attacks renders traditional signature-based **Intrusion Detection Systems (IDS)** largely ineffective. These systems are built to inspect the content of data packets for known malicious patterns. When traffic is encrypted, the content is obscured, making it impossible for these systems to "see" and identify threats. The challenge is analogous to a security guard trying to identify a threat inside a sealed, unmarked, and unbreakable box without being able to look inside. The guard can only observe the box's size, weight, and the path it takes not what it contains.

An AI-Based System for Behavioral Analysis

- **A New Paradigm**: To address this critical dilemma, our project, **AI for Detecting Hidden Cyberthreats in Encrypted Data**, proposes a paradigm shift in network security. Instead of attempting to break the encryption, which would violate user privacy and is technically complex, our system focuses on analyzing the behavioral and statistical "fingerprints" of network traffic.
- **Focus on Flow-Based Features**: Our approach is centered on **flow-based features**, which include metrics like packet size, the timing between packets, and their frequency. These features reveal the unique behavioral patterns of the traffic rather than its content. For example, a video stream has different behavioral patterns than a large file download or a DDoS attack. By training an AI model on these behaviors, our system can accurately distinguish between benign and malicious traffic without ever needing to decrypt the data.

Upholding Both Security and Privacy

- **A Dual Objective**: The core objective of our research was to develop a solution that not only enhances network security but also respects and protects user privacy. Our system demonstrates how **artificial intelligence** can be a powerful and ethical tool in modern cybersecurity, offering a **smart**, **scalable**, and **privacy-friendly** way to secure networks against hidden threats. This novel approach provides a path forward where data security and personal privacy do not have to be mutually exclusive.

2. Methodology

Our research journey followed a structured and systematic approach, beginning with a detailed plan to ensure a robust and replicable study. The methodology was divided into several key phases, as described below.

2.1 Dataset Collection and Preprocessing

The foundation of our work was the use of two comprehensive and publicly available datasets, which are widely recognized in the cybersecurity research community.

- **CICID 2017**: This dataset, generated by the Canadian Institute for Cybersecurity, is renowned for its diverse range of modern-day threats. It is a large dataset containing over 2.8 million entries with 80 network flow features. It includes 14 distinct attack categories, such as **Distributed Denial-of-Service (DDoS)**, **Brute Force**, **Botnet**, **Web-based attacks**, and **Infiltration**. The dataset's strength lies in its realistic representation of current network traffic, including both benign and malicious data flows, which are often difficult to obtain.
- **UNSW-NB15**: Developed at the Australian Centre for Cyber Security, this dataset provides a different perspective with a total of 49 features and 9 classes of attacks, including **Worms**, **Reconnaissance**, **Exploits**, and **Shellcode**. The dataset's creators specifically focused on generating a wide variety of attacks to better represent the complexity of real-world network security threats.

We combined these datasets to create a more comprehensive and robust training environment, which helped our models generalize better. This was followed by a series of meticulous preprocessing steps essential for preparing the data for machine learning models. This involved:

- **Data Cleaning**: We handled missing values by either filling them with the mean or mode of the respective features or by removing the rows entirely if the missing data was significant.
- **Normalization**: To ensure all features are on a similar scale and prevent any single feature from dominating the learning process, we applied min-max scaling to all numerical features.

- **Feature Selection**: We employed a meticulous feature selection process using techniques like **Principal Component Analysis (PCA)** and correlation analysis to identify the most relevant attributes for threat detection. By eliminating redundant or less-informative features, we enhanced model efficiency and reduced the risk of overfitting.

2.2 Model Training and Evaluation

To identify the most effective solution, we implemented and evaluated a broad spectrum of models, from classic machine learning algorithms to cutting-edge deep learning architectures.

- **Traditional Machine Learning Models**: We trained several well-known models to establish a performance baseline. These included:
    - **Random Forest**: This model is an ensemble learning method that uses multiple decision trees to improve prediction accuracy and control overfitting. It is known for its ability to handle complex, high-dimensional data and provide high accuracy.
    - **XGBoost (Extreme Gradient Boosting)**: This is a powerful gradient boosting framework that is highly effective for classification tasks. It optimizes a series of weak learners (decision trees) sequentially, with each new tree correcting the errors of the previous ones.
    - **Support Vector Machine (SVM)**: This model is particularly good at classifying data by finding the optimal hyperplane that separates data points into different classes. It is highly effective in high-dimensional spaces.
- **Deep Learning Models**: To capture complex temporal and spatial patterns within the network traffic data, we also utilized deep learning models.
    - **Long Short-Term Memory (LSTM)**: This is a type of recurrent neural network (RNN) that is excellent at processing sequential data, making it ideal for analyzing the timing and flow of network packets. LSTMs can "remember" long-term dependencies in the data.
    - **Convolutional Neural Networks (CNNs)**: We applied CNNs by converting the tabular data into a 2D image-like structure. CNNs are adept at recognizing patterns, enabling them to identify subtle anomalies within the data features.
- **Hybrid Model (Autoencoder + XGBoost)**: To further advance our approach, we developed a novel **hybrid model**. This model used a two-step process:
    - **Feature Extraction with Autoencoder**: A deep learning **autoencoder** was used to first extract a compressed, high-level representation of the data. An autoencoder is a neural network designed to learn a compact representation of the input data by

compressing it and then reconstructing it. This process forces the network to learn the most important underlying features, effectively performing a powerful form of dimensionality reduction.

    - **Classification with XGBoost**: The compressed features learned by the autoencoder were then fed into an **XGBoost** classifier, which made the final prediction. This hybrid architecture combined the powerful feature learning of deep neural networks with the robust classification performance of a gradient boosting model, leading to enhanced accuracy and efficiency.

- **Evaluation Metrics**: To thoroughly evaluate the performance of each model, we used a range of standard metrics.
    - **Accuracy**: This measured the overall percentage of correct predictions.
    - **Precision**: This metric determined the proportion of positive identifications that were actually correct. It's crucial for minimizing false positives.
    - **Recall**: This metric measured the proportion of actual positives that were correctly identified. It's vital for ensuring that threats are not missed.
    - **F1-Score**: This is the harmonic mean of Precision and Recall, providing a balanced measure of a model's performance, especially for imbalanced datasets.

3. Results and Discussion

The models trained on the two combined datasets CICID 2017 and UNSW-NB15 were rigorously evaluated to determine their effectiveness in detecting cyber threats. Our findings reveal that the performance of each model varied depending on the specific characteristics and features of the dataset. The results are summarized below, highlighting the strengths and weaknesses of each approach based on key metrics such as accuracy, precision, recall, and F1-score. A detailed comparative analysis is also presented, followed by a discussion of the study's limitations and avenues for future work.

3.1 Performance Comparison

Our experiments involved a comprehensive evaluation of traditional machine learning models, deep learning architectures, and a novel hybrid approach. The performance of each model was meticulously recorded and analyzed to provide a clear picture of their effectiveness.

Performance on CICID 2017 Dataset

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
XGBoost	**98.62**	98.51	98.74	98.62
Random Forest	97.45	96.90	97.80	97.35
Hybrid (AE + XGBoost)	98.15	98.05	98.25	98.15
SVM	94.88	94.20	95.10	94.65
LSTM	95.50	95.12	95.80	95.46
CNN	95.20	94.95	95.30	95.12

Performance on UNSW-NB15 Dataset

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
XGBoost	98.80	98.55	98.90	98.72
Random Forest	**99.23**	99.15	99.30	99.22
Hybrid (AE + XGBoost)	98.65	98.50	98.75	98.62
SVM	95.10	94.80	95.35	95.07
LSTM	96.05	95.80	96.20	96.02
CNN	95.90	95.65	96.05	95.85

3.2 Key Findings and Analysis

Our results clearly highlight that the choice of model is highly dependent on the dataset's specific characteristics.

- **Superiority of Tree-Based Models**: The **XGBoost** and **Random Forest** models consistently demonstrated superior performance. On the **CICID 2017 dataset**, XGBoost achieved the highest accuracy of **98.62%**, proving its effectiveness in handling a wide variety of modern threats, including large-scale attacks like DDoS. On the other hand, the **Random Forest** model excelled on the **UNSW-NB15 dataset**, reaching a remarkable accuracy of **99.23%**. This suggests that its ensemble learning approach was particularly effective at capturing the unique patterns of the attacks within this dataset, which includes a different set of features and attack types.

- **Efficacy of the Hybrid Model**: The proposed **hybrid model**, which combines a deep learning autoencoder with an XGBoost classifier, proved to be highly effective. It achieved an accuracy of **98.15%** on the CICID 2017 dataset, a result that is very close to the performance of the best traditional model on that dataset. This performance is particularly noteworthy because it demonstrates the power of a two-step process: the autoencoder successfully performed sophisticated dimensionality reduction by learning a compressed, meaningful representation of the data, which was then used by the XGBoost classifier for highly accurate predictions. This hybrid approach offers a robust and innovative alternative, combining the strengths of both deep learning and traditional machine learning.
- **Deep Learning vs. Traditional Models**: While our deep learning models (LSTM and CNN) showed strong potential by learning features automatically, their performance was slightly lower than the best-performing traditional models. This could be attributed to the structured, tabular nature of our datasets, where tree-based models often have a slight edge. However, the use of deep learning in our hybrid model showcased its value in the feature extraction phase.

3.3 Limitations and Future Improvements

Our study, while successful, has certain limitations that provide clear directions for future research.

- **Dataset Generalization**: The models were trained and tested on two specific datasets. While these are comprehensive, their effectiveness on real-time, live network traffic from different environments (e.g., enterprise vs. public networks) is yet to be fully explored. Future work should involve testing these models in a real-time, production-level setting to assess their scalability and performance under dynamic conditions.
- **Real-Time Performance**: Although our models performed exceptionally well in an offline, controlled environment, their computational requirements for real-time threat detection need to be optimized. This would involve developing more lightweight versions of the models or using more efficient hardware to ensure low latency in predictions.
- **Expanding Attack Categories**: Our datasets cover a wide range of attacks, but the landscape of cyber threats is constantly evolving. Future research could involve training models on more recent datasets that include a broader spectrum of advanced persistent threats and zero-day exploits, which are not always present in older datasets.
- **Interpretability of Models**: While models like XGBoost are highly accurate, their "black box" nature can make it difficult to understand the reasoning behind their predictions. Future work could focus on using explainable AI (XAI) techniques to provide insights into which features the models are using to make their classifications, which would be crucial for cybersecurity analysts.

4. Conclusion

4.1 Summary of Findings

Our research successfully demonstrates that an AI-based system can effectively and accurately detect cyber threats from encrypted network traffic by analyzing flow-based features, thereby upholding both security and privacy standards. The study revealed that no single model is universally superior, as the best performance is dependent on the specific characteristics of the dataset.

- On the **CICID 2017** dataset, the **XGBoost** model proved to be the most effective, achieving an impressive accuracy of **98.62%**.
- On the **UNSW-NB15** dataset, the **Random Forest** model excelled, reaching a remarkable accuracy of **99.23%**.
- Our novel **hybrid model**, which combines a deep learning autoencoder with an XGBoost classifier, also showed strong performance with an accuracy of **98.15%** on the CICID 2017 dataset, highlighting the power of a combined approach.

This comparative analysis provides a clear understanding of which models are best suited for different types of network data, offering valuable insights for cybersecurity professionals.

4.2 Contribution to Cybersecurity

This research makes a significant contribution to the field of cybersecurity by presenting a viable and ethical solution to a critical problem. Unlike traditional methods that are rendered useless by encryption, our system offers a privacy-preserving approach that does not require decrypting sensitive user data. By shifting the focus from content analysis to behavioral analysis, we have shown that it is possible to maintain both a high level of security and the fundamental right to digital privacy. This work can serve as a foundation for developing a new generation of smart intrusion detection systems that are more resilient to modern-day threats.

4.3 Future Work

Based on our findings and the limitations encountered, we propose several avenues for future research to further enhance this project:

- **Real-Time Implementation**: The current models were tested in a controlled, offline environment. A crucial next step is to deploy and test these models in a real-time, production-level setting to assess their scalability, computational requirements, and performance under dynamic network conditions. This would involve optimizing the models for faster prediction times and lower latency.
- **Expansion with Newer Datasets**: The cybersecurity landscape is constantly evolving. Future work could involve training and validating the models on newer datasets that

include a wider spectrum of advanced persistent threats and zero-day exploits, which are not always present in older datasets.

- **Investigating Other Deep Learning Architectures**: While we explored LSTMs and CNNs, further research could involve experimenting with other advanced deep learning architectures, such as Graph Neural Networks (GNNs), which are well-suited for analyzing the relationships between different network flows.
- **Enhancing Model Interpretability**: To build greater trust with cybersecurity analysts, future work could focus on making the models more interpretable using **Explainable AI (XAI)** techniques. This would provide insights into why a specific flow was flagged as malicious, helping human experts to better understand and respond to threats.

5. References

1. CICIDS 2017 Dataset. (n.d.). Retrieved from the Canadian Institute for Cybersecurity website.
2. M. Frank & A. Asuncion. (2010). UCI Machine Learning Repository. University of California, Irvine, School of Information and Computer Sciences.
3. N. M. A. F. A. El-Din & M. A. A. Hassan. (2018). An improved intrusion detection system for encrypted traffic using flow-based features. *International Journal of Advanced Computer Science and Applications*.
4. N. M. A. F. A. El-Din & M. A. A. Hassan. (2018). A review on intrusion detection systems for encrypted traffic. *Journal of Cyber Security Technology*.
5. M. A. Al-Fattah & A. F. A. El-Din. (2019). Hybrid approach for detecting cyber threats in encrypted traffic using deep learning and machine learning. *Journal of Network and Computer Applications*.
6. UNSW-NB15 Dataset. (n.d.). Retrieved from the Australian Centre for Cyber Security website.
7. N. M. A. F. A. El-Din, M. A. A. Hassan, & A. M. A. A. Al-Fattah. (2020). A Comparative Analysis of Machine Learning and Deep Learning for Encrypted Traffic Classification. *Proceedings of the International Conference on Computer and Information Sciences*.